

## APPARATUS AND METHOD FOR FILE CAPTURE, PRESERVATION AND MANAGEMENT

[0001] This application claims the benefit of Provisional Application No. 60/234,221 filed September 21, 2000, which is herein incorporated by reference.

### Field of the Invention

[0002] This invention relates to file preservation, file capture, file management and file integrity techniques for computing devices.

### Background of the Invention

[0003] One of the greatest challenges faced by information technology (IT) professionals and computer users today, particularly in the business environment is the protection and management of data. Data may be stored on user workstations, e.g., laptop computers, home or office desktop computers, network servers or other devices external to the workstations. Important data may even be stored on hand-held computing devices such as PDAs, PALs and other like devices. Complicating the problem is the fact that the criticality of data is increasing and the difficulty of managing it, protecting it from loss and keeping it available is increasing. This is due to a variety of factors, including: 1) the explosion in data volume, particularly that stored on desktop and laptop computers, 2) the increasing complexity of desktop and laptop computer software and hardware and increasing trends toward a paperless environment where absolute reliance (because paper copies are becoming less the norm) on data integrity is increasingly significant.

[0004] Many home computer users do not realize the vulnerability of their computer data. Many that do understand the very real potential for data loss, purchase backup

systems whose operation and user interface is often confusing and/or time-consuming to use, dramatically decreasing their effectiveness or dependability. As a result, many computer users remain very much at risk of data loss resulting from hardware and/or software failures, fires, stolen equipment, etc. While these risks are significant, the most frequent cause of data loss is user error (accidental file deletes, file overwrites, errant programs, etc.), to which users remain very vulnerable even with most present day backup systems.

**[0005]** The financial impact of information loss is substantial. As reported by the Safeware Insurance Agency, in 1999 alone, insurance claims for damaged, lost and stolen computers (primarily notebook computers) totaled more than \$1.9 billion. This figure does not include the untold billions lost in intellectual capital and time. It is costly to recreate lost data and there are significant related costs such as lost productivity and lost opportunity. Consider, for example, the financial and health related impact of a doctor losing all patient contact information and medical histories due to a hard disk crash or some other type of computer failure. In addition, it is costly to keep desktop and laptop computers up and running in the wake of their increasing complexity.

**[0006]** A variety of products have been developed to address data preservation and integrity issues. These products may be loosely grouped into three categories, manual backup systems, schedule based backup systems and mirroring backup systems.

**[0007]** The least efficient and probably one of the most frequently used backup systems is the manual backup. At times determined by the user, the user selects files to be backed up and either utilizes the built in backup procedure for the corresponding application or manually copies the selected files to a desired backup storage media.

**[0008]** The problems with this method of preserving data are self-evident. Backup procedures are often confusing and may differ from application to application. Accordingly, the user must familiarize itself with the various methods for performing backups. In addition, users may forget to backup or elect not to on a given occasion due to time constraints or other reasons. Manual backups often do not allow the user to continue to use the system during the backup procedure. Furthermore, data stored to the backup media is really only a “snapshot” of the data at the time that the backup is performed. Any changes made between manual backups would be lost if there was a failure on the computer’s storage device.

**[0009]** Schedule based backup systems typically perform backups according to a schedule either set by the user or preset by the backup software. One of the major disadvantages of each type of schedule-based backup system is that as with manual backups, they miss work done between schedule points. This may cause the user to lose critical information as they work between schedule points. Another disadvantage of schedule-based backups systems is that they are frequently confusing and cumbersome for the user. Still another disadvantage of schedule-based backup systems is that they function poorly if at all when the backup storage device is unavailable, i.e., they cannot be written to due to a communications error or because the device has reached its capacity, is bandwidth limited, or is non-operational for some other reason.

**[0010]** Mirroring is a technique typically applied to disk based backup systems. Mirroring backup systems are the most comprehensive in that everything that happens to the source storage device immediately happens to the backup storage device. That

is the backup drive becomes a mirror image of the source drive. Accordingly, if a failure occurs on the source disk, processing can be switched to the backup disk with little or no service interruption.

**[0011]** The strongest advantage of mirroring systems is also their strongest disadvantage. Because there is no operational discrimination, if a file is accidentally deleted from the source disk, it is deleted and cannot be preserved on the backup disk. Likewise, if a virus infects the source disk it is likely to infect the backup disk. Another disadvantage of mirroring systems is that separate backup disks are required for each source disk, doubling the disk requirement for the system. The backup disk must be at least as large as the source disk and the disks must be configured with identical volume mapping. Any extra space that may be present on the backup disk is unavailable.

**[0012]** All of these methods require that the user specify which files/directories to back up, but many users have no concept of files and directories in their thought process, much less are they able to correlate a particular application (e.g. Microsoft Excel) with the kinds and locations of files they generate. These systems simply require too much user knowledge, and too much user intervention. The backup user's risk increases dramatically the lower his computer knowledge may be.

**[0013]** In view of the foregoing, there is a need for a file capture, preservation and management system that captures files just before and/or just after they have been changed to minimize loss of data between backup events. There is also a need for file capture and preservation system that captures files even when the destination storage medium for the files is unavailable. There is a further need for a system that allows users to recover easily and quickly from any type of information loss, including simple

user errors, failed software installations or updates, hardware failures (attached storage devices), and lost or stolen laptop computers. Users should be able to recover on their own, without the intervention of the IT staff, and their backup systems should be as “behind the scenes” as possible, requiring little user attention and extremely small amounts of user computer knowledge.

### **Summary of the Invention**

**[0014]** It is an object of the invention to a file capture, preservation and management method and apparatus that captures files just before and/or just after the files are changed.

**[0015]** It is another object of the invention to provide a file capture, preservation and management method and apparatus that has an imperceptible impact on system performance from the user's point of view.

**[0016]** It is a further object of the invention to provide a file capture, preservation and management method and apparatus that captures and stores files even when there is no connection to the desired storage location.

**[0017]** Still another object of the invention is to provide a file capture, preservation and management method and apparatus that captures and stores files even when the desired storage location is unavailable.

**[0018]** In accordance with an aspect of the invention, a method for archiving files is provided. The method includes, in a computing device, detecting an instruction from a resident program to perform an operation on an operating file. Upon detection of the instruction, capturing the operating file temporally proximate to the operation being performed on the operating file.

**[0019]** In accordance with another aspect of the invention, a method for moving files from a first storage location to a second storage location is provided. The method includes, in a computing device, searching a first storage location for files responsive to the occurrence of a first event and moving the files from the first storage location to the second storage location responsive to a second event.

**[0020]** In accordance with still another aspect of the invention, a method for archiving files is provided. The method includes detecting an instruction from a resident program to perform an operation on an operating file. The method further includes creating an archive file from the operating file and storing the archive file in a first storage location temporally proximate to the operation being performed on the operating file and responsive to detecting the instruction. In keeping with the method, the first storage location is searched for an archive file responsive to the occurrence of a first event. The archive file is then moved from the first storage location to the second storage location responsive to a second event.

**[0021]** The accompanying figures show illustrative embodiments of the invention from which these and other of the objectives, novel features and advantages will be readily apparent.

#### **Brief Description of the Drawings**

**[0022]** Figure 1 illustrates a block diagram of a computing device in accordance with the present invention.

**[0023]** Figure 2A is a flow chart depicting a process for moving files in accordance with the present invention.

[0024] Figure 2B is a flow chart showing another process for moving files in accordance with the invention.

[0025] Figure 3 is a time line illustrating a sequence of events in an exemplary operation in accordance with the invention.

[0026] Figure 4 is a time line illustrating a sequence of events in another exemplary operation in accordance with the invention.

[0027] Figure 5 is a time line illustrating a sequence of events in still another exemplary operation in accordance with the invention.

### **Detailed Description of the Embodiment**

#### **DEFINITIONS**

[0028] **Operating System (OS)** – A computer program that allocates system resources such as memory, disk space, and processor usage and makes it possible for the computer to boot up to a human user interface allowing the user to interact with the computer and control its operation.

[0029] **Operating File** – a system or user file.

[0030] **Archive File** – a file containing all of the data of an operating file in a native or altered format and/or a file containing at least some of the data of an operating file and including references to the location of the remainder of the data of the operating file.

[0031] **Computing Device** – a personal computer, a laptop or notebook computer, a server, a hand-held computing device, a PDA or a PAL. The term computing device is not specific to the kind of operating system being run on such computing device, and includes devices running Microsoft operating systems, Apple Macintosh operating

systems, UNIX operating systems, Linux operating systems, and other operating systems.

**[0032] Storage Location** – any storage device, or a buffer, folder, directory or designated area on a storage device.

**[0033] Personal Attached Storage Device** -- any internal or external storage device connected to a computing device.

**[0034] Network Attached Storage Device** -- any storage device connected directly to a network to which a first computing device is also temporarily or permanently connected, or any storage device connected to a second computing device that is also temporarily or permanently connected to the network to which the first computing device is temporarily or permanently connected.

**[0035] Internet storage area network** -- any storage area (device, collection of devices, etc.) that can be accessed by the computing device when the computing device is temporarily or permanently connected to the Internet.

**[0036] Peer-to-Peer Storage Device** -- any storage area (device, collection of devices, etc.) that can be accessed by the computing device when it is sharing resources with other network or Internet accessible computers.

**[0037] Resident Program** – an operating system (OS) or other program that has control over file operations such as “read”, “write”, “save”, “rename”, “delete”, “copy”, “move”, “open”, “close”, etc.

**[0038] User Program** – an application software program or other computing program installed by the user or by the computer manufacturer for user creation of desired data, documents, or other information that is designed to enhance the

functionality and/or enjoyment and/or usability of the computing device. The present invention is directed to an apparatus and/or method for file capture, preservation and management. The invention includes a file capture aspect and smart data management aspect. The invention may be realized as a method and/or an apparatus. More particularly, the invention may be realized as a set of program code instructions stored on a computer usable medium, a set of program code instructions embodied in a signal for transmitting computer information, and a processor and/or computing device configured as described herein.

**[0039]** Figure 1 depicts a block diagram in accordance with the present invention comprising a computing device 5 including a file capture block 10 (or file capturer), a smart data management block 15 (or smart data manager), an input buffer 20, output buffer(s) 25, and a database 30. A storage device 35 is also provided and may be either internal or external to computing device 5. The invention functions in conjunction with a resident program on computing device 5.

**[0040]** In accordance with an embodiment of the invention, file capture block 10 detects an instruction to perform an operation on an operating file initiated by the resident program of computing device 5. At a moment temporally proximate to when the resident program actually performs the operation, i.e., just before and/or just after the operation is performed on the operating file, or, more preferably, the instant before and/or the instant after the operating file is changed, file capture block 10 captures the operating file or portions thereof. Preferably, the operating file is captured within a few clock cycles of the detection of the instruction.

**[0041]** In keeping with a preferred aspect of the invention, file capture block 10 causes the location of the captured operating file to be recorded in database 30. The continued process of recording information about captured operating files, or portions thereof, in database 30 creates a record of each version of the operating file, which may be accessed by the user or by other programs.

**[0042]** File capture is preferably executed by creating an archive file from the operating file. The archive file is preferably stored in a temporary storage location, internal or external to the computer, such as input buffer 20. However, the archive file may be stored directly in storage device 35. In accordance with a preferred aspect of the invention, storage device 35 may be a personal attached storage device, a network attached storage device, an Internet storage area network, a peer-to-peer storage device, or other storage device.

**[0043]** In keeping with a preferred aspect of the invention, smart data management block 15 manages the migration of the archive file from the input buffer 20 through the output buffers 25 to storage device 35. This migration may take place either synchronously or asynchronously with the file capture procedures described herein. The time duration from a file arriving in input buffer 20 and when it arrives on archive storage device 35 is managed by the smart data management block 15. More particularly smart data management block 15 regularly examines input buffer 20 for the presence of archive files. Smart data management block 15 performs this examination upon the occurrence of an event, e.g., messages from the file capture block 10 and/or various messages from the resident program(s), messages from an input buffer timer sent at time intervals controlled by a timer or at time intervals selected by the user.

Optionally, smart data management block 15 may then examine database 30 to determine a defined storage location for each of the archive files stored in input buffer 20. Each archive file stored in the input buffer 20 may be directed to the same storage location or to different storage locations and archive files may be directed to multiple storage locations for redundancy. Preferably, smart data management block 15 moves the archive files to one or more output buffers 25. More preferably each archive file is moved to output buffer(s) 25 corresponding to the final storage location(s) for that archive file. Alternatively, all archive files may be moved to a single common output buffer 25 if desired. Upon the occurrence of an event, and/or at defined time intervals, smart data management block 15 moves the archive files from the output buffers 25 to their respective storage device(s) 35. Exemplary events include but are not limited to messages indicating when storage device 35 is connected and ready for use, messages indicating when storage device 35 is inserted/removed, full, defective, etc., and messages indicating when storage device 35 is disconnected or unavailable, and messages from a storage device timer sent at time intervals controlled by the timer or at time intervals controlled by the user. The input buffer timer and the storage device timer may operate synchronously or non-synchronously.

**[0044]** Under certain conditions, smart data management block 15 may be unable, or may elect not to move the archive files. For example, if storage device 35 is unavailable then smart data management block 15 will not move the archive files to storage device 35. Among the conditions that may cause storage device 35 to be unavailable are i) storage device 35 is disconnected from computing device 5, ii) the connection between storage device 35 and computing device 5 is faulty or unacceptably

slow, iii) storage device 35 is full, or iv) storage device 35 is malfunctioning. In addition, smart data management block 15 may also regulate movement of archive files according to time schedules set by the user, by monitoring connection bandwidth availability and moving files only during times of high bandwidth availability, or by monitoring other factors including messages that may received from storage location server requests for archive file transmittal.

**[0045]** A preferred operational mode for smart data management block 15 is illustrated in the flowcharts of Figures 2A and 2B. In step 100 of Figure 2A, smart data manager 15 examines input buffer 20 to determine whether any archive files are stored therein. If no archive files are present, smart data manager 15 rests idle until the next event occurs. If archive files are detected, in step 105, smart data manager 15 updates database 25 to indicate the location of the archive files; that is, to indicate that the archive files are resident in input buffer 20. In step 110, smart data manager 15 examines database 30 to determine the proper destination for each archive file. In step 115, smart data manager 15 moves the archive files to output buffers 25. In step 120, smart data manager 15 updates database 30 to indicate that the archive files are now stored in the output buffer.

**[0046]** In Figure 2B in step 125, the archive files are moved to one or more storage devices 30. If smart data manager 15 is unable to move the archive files to any of the storage devices 30, smart data manager 15 rests idle and does not move the archive files until it is notified that the storage device is available. Accordingly, the archive files remain in either input buffer 20 or output buffer 25 until smart data management block

15 is notified. In step 130 smart data manager 15 updates database 25 to indicate that the archive files are stored in one or more storage devices 30.

## **Use Specific to User Program Operations**

**[0047]** The following examples are directed to embodiments of the invention specific to operations performed by a user program. The file capture, preservation and management processes of the invention are not limited to execution with the exemplary operation discussed below. The processes of the invention are preferably executed when a resident program causes a change or a change to be imminent in the operating file. Therefore, the following examples are intended to be exemplary only and non-limiting.

### **File Capture at File Open**

**[0048]** As illustrated in Figure 3, in step 205, the user or a program selects an “open” operation to open an operating file and an instruction to perform that “open” operation on the operating file is sent to the resident program. In step 210, file capture block 10 detects the instruction and captures the operating file. Optionally, prior to capturing the operating file, file capture block 10 may check database 30 to a) determine whether the operating file has previously been archived, b) determine whether the user has selected the operating file for protection, or c) determine a match to other defined conditions. If the go-ahead conditions exist, then file capture block 10 creates an archive file and stores the archive file in a storage location such as input buffer 20 or storage device 35 just before the resident program opens the operating file. Preferably, file capture block 10 stores the archive file in input buffer 20. In step 215 the resident program opens the operating file and in step 220 the user program displays the

operating file as originally requested, e.g. Microsoft Word, and makes it available for the user to alter, e.g., edit a word processing document, amend or add to a database, etc. Step 210 is performed by momentarily delaying the execution of step 215 in such a manner as to have little or no perceptible impact on system performance from the user's point of view.

**[0049]** In step 225, the user program begins a process to save the altered operating file and an instruction to save the altered operating file is sent to the resident program. In step 230 the resident program saves the altered operating file pursuant to the instruction. In step 235, immediately after the altered operating file is saved by the resident program, file capture block 10 captures the altered operating file, preferably by creating and storing an archive file of the altered operating file in input buffer 20. In accordance with a preferred feature of the invention, file capture block 10 may save the archive file in such a way that previous revisions of the operating file are retained. That is, every time the operating file is changed, file capture block 10 saves an archive file and database 30 is updated with information about the archive file. Accordingly, over time, a plurality of archive files may be created from the original operating file. Each archive file represents a revision of the original operating file.

#### **File Capture in the "RENAME" Operation**

**[0050]** As illustrated in Figure 4, step 305, in performing an operating file rename operation, the user or a program generates an instruction for the resident program to select a new name for an old operating file. In step 310, file capture block 10 detects the instruction and captures the old operating file. Optionally, prior to capturing the old operating file, file capture block 10 may check database 30 to a) determine whether the operating file has previously been archived, b) determine whether the user has selected

the operating file for protection, or c) determine a match to other defined conditions. If the go-ahead conditions exist, then file capture block 10 creates an archive file of the old operating file and stores the archive file in a storage location such as storage device 35 or, more preferably, input buffer 20 just before the resident program renames the old operating file. In step 315 the resident program renames the old operating file, thus creating a new operating file. Immediately after the old operating file is renamed, file capture block 10 captures the new operating file. Optionally, prior to capturing the new operating file, file capture block 10 may determine whether the new operating file has previously been archived, whether the user has selected the new operating file for protection, or other matching conditions exist. Like the archive file for the old operating file, the archive file for the new operating file is preferably stored in input buffer 20. In step 325 file capture block 10 and smart data management block 15 associate or link the new operating file with each of the versions of the old operating file to create a continuous operating file revision history.

#### **File Capture in the “Delete” Operation**

**[0051]** Figure 5 illustrates the file capture process in the delete operation. In step 405, the user or a program identifies an operating file to delete and generates an instruction to the resident program. In step 410, file capture block 10 detects the instruction and captures the operating file just before it is deleted in step 415. Optionally, prior to capturing the operating file, file capture block 10 may check database 30 to a) determine whether the operating file has previously been archived, b) determine whether the user has selected the operating file for protection, or c) determine a match to other defined conditions. If the go-ahead conditions exist, then file

capture block 10 preferably captures the operating file. In step 420, the resident program deletes the operating file.

**[0052]** As shown by the examples given, a clear advantage of the invention is, regardless of the operation being performed, after each file capture step, file capture block 10 preferably updates database 30 to indicate the location of the corresponding archive file. Database 30 may keep track of multiple versions of an operating file, any of which may be accessed at the request of the user or other program.

**[0053]** Another advantage of the invention is that by capturing the operating file just before and/or just after an operation is performed thereon, the invention achieves near real-time operating file archiving while achieving minimal missed alterations to an operating file.

**[0054]** A further advantage of the invention in its preferred embodiment, is that by intelligently managing the migration of operating files from the input buffer 20 through the output buffer 25 to the storage device 35, the invention achieves protection of operating files even when the desired storage device is permanently or temporarily unavailable.

### Industrial Applicability

**[0055]** The present invention is suited for any application that requires or benefits from near real time file capture, that seeks improved file integrity and/or that seeks efficient management of file storage. For example, the present invention is particularly useful in backup systems, audit trail systems, computer security systems, systems for monitoring computer users and others.

**[0056]** Although the present invention has been described in terms of particular preferred embodiments, it is not limited to those embodiments. Alternative embodiments, examples, and modifications which would still be encompassed by the invention may be made by those skilled in the art, particularly in light of the foregoing teachings.